Sekretarov problem - Maksimizacija očekivanog kvaliteta kandidata

30. novembar, 2018

Apstrakt

Predstavljamo vam modifikaciju Sekretarovog problema u kojoj sekretar bira tačno jednog od n kandidata čiji su kvaliteti $1, 2, ..., n$. Kandidati na intervju dolaze u nasumično poretku. Sekretar ne može da intervjuiše istu osobu više puta. Intervjuisanu osobu može da poredi isključivo sa prethodnim kandidatima. Njegov cilj je da prosečan kvalitet izabranog kandidata bude što veći. Dokazali smo da se optimalna strategija sastoji od odbijanja prvih $\sqrt{n} - 1$ kandidata, a onda biranja prvog sledećeg koji je bolji od svih prethodnih.

1 Uvod

Pravila Sekretarovog problema su:

1. Postoji sekretar koji želi da zaposli jednu osobu
2. Znamo koliko se ljudi prijavilo za posao. Neka je taj broj n
3. Svi kandidati su različitog kvaliteta
4. Kandidati su intervjuisani u nasumičnom poretku
5. Nakon što ispitamo osobu moramo je ili odmah primiti ili odbiti
6. Ako osobu odbijemo više je nikada nećemo videti. Ako je prihvatimo više nikoga ne ispitujemo
7. Naša odluka da li ćemo prihvatiti osobu je napravljena isključivo poređenjem tog kandidata sa prethodno ispitanim osobama.

$\boxed{8.}$ Naš cilj je da napravimo strategiju takvu da su nam šanse za biranje najboljeg kandidata najveće.

Mi smo modifikovali ta pravila i napravili varijaciju Sekretarovog problema koju smo nazvali Maksimizacija očekivanog kvaliteta izabranog kandidata. Njena pravila su:

1. Postoji sekretar koji želi da zaposli jednu osobu
2. Znamo koliko se ljudi prijavilo za posao. Neka je taj broj n
3. Svi kandidati su različitog kvaliteta
4. Kandidati su intervjuisani u nasumičnom poretku
5. Nakon što ispitamo osobu moramo je ili odmah primiti ili odbiti
6. Ako osobu odbijemo više je nikada nećemo videti. Ako je prihvatimo više nikoga ne ispitujemo
7. Naša odluka da li ćemo prihvatiti osobu je napravljena isključivo poređenjem tog kandidata sa prethodno ispitanim osobama.

$\boxed{8.}$ Naš cilj je da napravimo strategiju takvu da nam je prosečan kvalitet izabranog kandidata što veći

$\boxed{9.}$ Kvaliteti kandidata su od 1 do n gde je n najbolji kvalitet.

2 Optimalna strategija i cilj

Nađena je optimalna strategija koje sekretar treba da se drži. On prvo odbija određeni broj kandidata nezavisno od njihovog kvaliteta a nakon toga bira prvog sledećeg koji je bolji od svih prethodnih. Ako takva osoba ne postoji onda je sekretar primoran da izabere poslednjeg kandidata. Cilj je da nađemo zavisnost broja kandidata koje odbijemo u odnosu na ukupan broj kandidata tako da nam prosečni kvalitet izabranog kandidata bude najveći moguć.

3 Dokaz

3.1 $P(n,k,x)$

$P(n,k,x)$ je verovatnoća da izaberemo kandidata kvaliteta x ako od n odbijemo prvih k kandidata. Verovatnoća da to uradimo je jednaka verovatnoći da se kandidat kvaliteta x nalazi na prvom mestu i da ga izaberemo ako se nalazi na prvom mestu plus verovatnoća da se kandidat kvaliteta x nalazi na drugom mestu i da ga izaberemo ako se nalazi na drugom mestu plus ... plus verovatnoća da se kandidat kvaliteta x nalazi na n-tom mestu i da ga izaberemo ako se nalazi na n-tom mestu. To možemo matematički zapisati kao:
$P(n,k,x) = p(x \text{ na mestu broj } 1) \cdot p(\text{ izabran } |x \text{ na mestu broj } 1) + p(x \text{ na mestu broj } 2) \cdot p(\text{ izabran } |x \text{ na mestu broj } 2) + ... + p(x \text{ na mestu broj } n) \cdot p(\text{ izabran } |x \text{ na mestu broj } n)$
Pošto smo odbili prvih k kandidata:

$\sum_{i=1}^{k} p(x \text{ na mestu broj } i) \cdot p(\text{ izabran } |x \text{ na mestu broj } i) = \sum_{i=1}^{k} \frac{1}{n} \cdot 0 = 0$
Primetimo da je $p(\text{ izabran } |x \text{ na mestu broj } i) = 0$ jer je nemoguće izabrati kandidata na mestu i ako za i važi : $1 \leq i \leq k$ jer odbijamo prvih k kandidata. Dakle znamo da važi:
$P(n,k,x) = \sum_{i=k+1}^{n} p(x \text{ na mestu broj } i) \cdot p(\text{ izabran } |x \text{ na mestu broj } i)$

Kandidata kvaliteta $x < n$ možemo izabrati u 2 slučaja:

1) kada je najbolji do sada

2) kada nije najbolji do sada ali je poslednji pa smo primorani da ga izaberemo

Kandidata kvaliteta $x = n$ možemo izabrati isključivo u slučaju kada je on najbolji do sada, jer čak i ako se nalazi na poslednjem mestu i dalje je sigurno najbolji do sada pa ga nismo izabrali jer smo primorani već iz razloga što nam to odgovara.
Da bismo izabrali kandidata kvaliteta x na mestu i kao najboljeg do sada kandidati pre njega moraju imati manji kvalitet. Tih slučajeva ima $\binom{x-1}{i-1}$. Kvaliteti ostalih kandidata mogu biti nasumični pa stoga imamo $(n-i)!$ permutacija. Takođe najkvalitetniji od prvih $i-1$ kandidata se mora nalaziti u prvih k jer bi u suprotnom njega izabrali pre kandidata kvaliteta x. To se dešava u $k(i-2)!$ slucajeva. Sve delimo ukupnim brojem permutacija $n!$ i dobijamo $\frac{\binom{x-1}{i-1}(n-i)!k(i-2)!}{n!}$. Ali, može se desiti da je kandidat izabran kao poslednji iako nije najbolji do sada. Verovatnoća da se kandidat kvaliteta x našao na poslednjem mestu jeste $\frac{1}{n}$. U ovom slučaju se najbolji kandidat našao u prvih k. Verovatnoća za tako nešto je: $\frac{k}{n-1}$. Dakle, verovatnoća da smo izabrali kandidata, iako nije najbolji, iz razloga što se našao na poslednjem mestu jeste: $\frac{k}{n(n-1)}$. Dakle,

$$P(n,k,x) = \begin{cases} \sum_{i=k+1}^{n} \frac{\binom{x-1}{i-1} \cdot k \cdot (i-2)! \cdot (n-i)!}{n!} + \frac{k}{n(n-1)} & \text{, za } n \neq x \\ \sum_{i=k+1}^{n} \frac{\binom{x-1}{i-1} \cdot k \cdot (i-2)! \cdot (n-i)!}{n!} & \text{, za } n = x \end{cases}$$

$$P(n,k,x) = \begin{cases} \sum_{i=k+1}^{x} \frac{(x-1)!\cdot k\cdot(i-2)!\cdot(n-i)!}{n!\cdot(i-1)!\cdot(x-i)!} + \frac{k}{n(n-1)} & \text{, za } n \neq x \\ \sum_{i=k+1}^{x} \frac{(x-1)!\cdot k\cdot(i-2)!\cdot(n-i)!}{n!\cdot(i-1)!\cdot(x-i)!} & \text{, za } n = x \end{cases}$$

$$\boxed{P(n,k,x) = \begin{cases} \sum_{i=k+1}^{x} \frac{(x-1)!\cdot k\cdot(n-i)!}{n!\cdot(i-1)\cdot(x-i)!} + \frac{k}{n(n-1)} & \text{, za } n \neq x \\ \sum_{i=k+1}^{x} \frac{(x-1)!\cdot k\cdot(n-i)!}{n!\cdot(i-1)\cdot(x-i)!} & \text{, za } n = x \end{cases}}$$

3.2 expValue(n, k)

expValue(n, k) je funkcija koja vraća očekivani kvalitet izabranog kandidata ako od n odbijemo prvih k. Nju računamo kao zbir proizvoda kvaliteta i verovatnoće da smo izabrali kandidata tog kvaliteta.

$$\boxed{\text{expValue}(n,k) = \sum_{j=1}^{n}(j \cdot P(n,k,j))}$$

3.3 maximizingExpectation(n)

maximizingExpectation(n) je funkcija koja vraća najveći očekivani kvalitet izabranog kandidata u zavisnosti od broja kandidata.

$$\boxed{\text{maximizingExpectation}(n) = \max(\text{expValue}(n,1),\ \text{expValue}(n,2), ..., \text{expValue}(n,n-1))}$$

3.4 Leme

Lema 3.1 $\boxed{\dfrac{1}{x(x+2)} = \dfrac{1}{2x} - \dfrac{1}{2(x+2)}}$

Dokaz:

$\frac{1}{2x} - \frac{1}{2(x+2)} = \frac{x+2}{2x(x+2)} - \frac{x}{2x(x+2)} = \frac{\cancel{x}+2-\cancel{x}}{2x(x+2)} = \frac{\cancel{2}}{\cancel{2}x(x+2)} = \frac{1}{x(x+2)}$

Lema 3.2 $\boxed{\displaystyle\sum_{i=2}^{n} \frac{1}{(i-1)(i+1)} = \frac{3n^2-n-2}{4n(n+1)}}$ *za* $n \in \mathbb{N}$

Dokaz:

Iz Leme1: $\frac{1}{(i-1)(i+1)} = \frac{1}{2(i-1)} - \frac{1}{2(i+1)}$

$\implies \sum_{i=2}^{n} \frac{1}{(i-1)(i+1)} = \sum_{i=2}^{n}(\frac{1}{2(i-1)} - \frac{1}{2(i+1)}) = \frac{1}{2} - \cancel{\frac{1}{6}} + \frac{1}{4} - \cancel{\frac{1}{8}} + \cancel{\frac{1}{6}} - \cancel{\frac{1}{10}} + \cancel{\frac{1}{8}} - \cancel{\frac{1}{12}} + ... + \cancel{\frac{1}{2(n-3)}} - \cancel{\frac{1}{2(n-1)}} + \cancel{\frac{1}{2(n-2)}} - \frac{1}{2n} + \cancel{\frac{1}{2(n-1)}} - \frac{1}{2(n+1)} = \frac{3}{4} - \frac{1}{2n} - \frac{1}{2(n+1)} = \frac{3n(n+1)-2(n+1)-2n}{4n(n+1)} = \frac{3n^2+3n-2n-2-2n}{4n(n+1)} = \frac{3n^2-n-2}{4n(n+1)}$

Lema 3.3 $\boxed{\displaystyle\sum_{i=k+1}^{n} \frac{1}{(i-1)(i+1)} = \frac{-k^2(2n+1)+k(2n^2-1)+n(n+1)}{2k(k+1)n(n+1)}}$

za $n \in \mathbb{N}$

Dokaz:

$\sum_{i=k+1}^{n} \frac{1}{(i-1)(i+1)} = \sum_{i=2}^{n} \frac{1}{(i-1)(i+1)} - \sum_{i=2}^{k} \frac{1}{(i-1)(i+1)}$

$= \frac{3n^2-n-2}{4n(n+1)} - \frac{3k^2-k-2}{4k(k+1)} = \frac{(3n^2-n-2)k(k+1)-(3k^2-k-2)n(n+1)}{4k(k+1)n(n+1)}$
$= \frac{\cancel{k^2\cdot 3n^2}+k\cdot 3n^2-k^2\cdot n-k\cdot n-k^2\cdot 2-k\cdot 2-\cancel{k^2\cdot 3n^2}-k^2\cdot 3n+k\cdot n^2+k\cdot n+2n^2+2n}{4k(k+1)n(n+1)}$
$= \frac{-k^2(n+2+3n)+k(3n^2-\cancel{n}-2+n^2+\cancel{n})+2n^2+2n}{4k(k+1)n(n+1)} = \frac{-k^2(4n+2)+k(4n^2-2)+2(n^2+n)}{4k(k+1)n(n+1)}$
$= \frac{2(-k^2(2n+1)+k(2n^2-1)+n(n+1))}{4k(k+1)n(n+1)} = \frac{-k^2(2n+1)+k(2n^2-1)+n(n+1)}{2k(k+1)n(n+1)}$

Lema 3.4 $\boxed{\sum_{i=1}^{n} i\binom{i-1}{t} = (t+1)\binom{n+1}{t+2}}$ *za* $t > 0,\ n, t \in \mathbb{N}$

Dokaz:
Indukcijska hipoteza:
$\sum_{i=1}^{1} i\binom{i-1}{t} = 1\binom{0}{t} = 0 = (t+1)\cdot 0 = (t+1)\binom{2}{t+2} = (t+1)\binom{1+1}{t+1}$ ✓
Baza indikcije:
$\sum_{i=1}^{n} i\binom{i-1}{t} = (t+1)\binom{n+1}{t+2}$
Indukcijski korak:
$\sum_{i=1}^{n+1} i\binom{i-1}{t} = \sum_{i=1}^{n} i\binom{i-1}{t} + (n+1)\binom{n}{t} = (t+1)\binom{n+1}{t+2} + (n+1)\binom{n}{t}$
$= \frac{(n+1)!(t+1)}{(n-t-1)!(t+2)!} + \frac{(n+1)}{(n-t)!t!} = (n+1)![\frac{(t+1)(n-t)}{(n-t)!(t+2)!} + \frac{(t+1)(t+2)}{(n-t)!(t+2)!}]$
$= (n+1)!\frac{t+1}{(n-t)!(t+2)!}[n - \cancel{t} + \cancel{t} + 2] = (t+1)\frac{(n+2)!}{(n-t)!(t+2)!} = (t+1)\binom{n+2}{t+2}$ ✓

3.5 Optimalno k

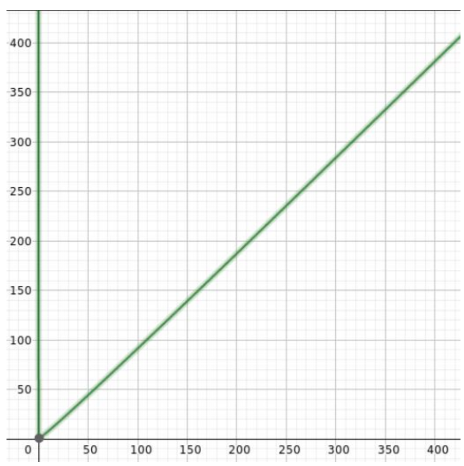
expValue(n,k) $= \sum_{j=1}^{n} j$P(n,k,j)
$= \sum_{j=1}^{n-1} j(\sum_{i=k+1}^{n-1} \frac{\binom{j-1}{i-1}k(i-2)!(n-i)!}{n!} + \frac{k}{n(n-1)}) + n\sum_{i=k+1}^{n} \frac{\binom{n-1}{i-1}k(i-2)!(n-i)!}{n!}$
$= \sum_{j=1}^{n-1} j(\sum_{i=k+1}^{n-1} \frac{\binom{j-1}{i-1}k(i-2)!(n-i)!}{n!}) + \sum_{j=1}^{n-1} \frac{jk}{n(n-1)} + n\sum_{i=k+1}^{n} \frac{\binom{n-1}{i-1}k(i-2)!(n-i)!}{n!}$
$= \sum_{j=1}^{n-1} j(\sum_{i=k+1}^{n} \frac{\binom{j-1}{i-1}k(i-2)!(n-i)!}{n!}) + n\sum_{i=k+1}^{n} \frac{\binom{n-1}{i-1}k(i-2)!(n-i)!}{n!} + \frac{k}{n(n-1)}\sum_{j=1}^{n-1} j$
$= \sum_{j=1}^{n} j(\sum_{i=k+1}^{n} \frac{\binom{j-1}{i-1}k(i-2)!(n-i)!}{n!}) + \frac{k}{\cancel{n(n-1)}} \cdot \frac{\cancel{(n-1)n}}{2}$
$= \sum_{j=1}^{n} \sum_{i=k+1}^{n} \frac{j\binom{j-1}{i-1}k(i-2)!(n-i)!}{n!} + \frac{k}{2}$
$= \sum_{i=k+1}^{n} \sum_{j=1}^{n} \frac{j\binom{j-1}{i-1}k(i-2)!(n-i)!}{n!} + \frac{k}{2}$
$= \sum_{i=k+1}^{n} \frac{k(i-2)!(n-i)!}{n!} \sum_{j=1}^{n} j\binom{j-1}{i-1} + \frac{k}{2}$
Iz Leme4: $\sum_{j=1}^{n} j\binom{j-1}{i-1} = i\binom{n+1}{i+1}$
$= \sum_{i=k+1}^{n} \frac{k(i-2)!(n-i)!}{n!} \sum_{j=1}^{n} j\binom{j-1}{i-1} + \frac{k}{2}$
$= \sum_{i=k+1}^{n} \frac{k(i-2)!(n-i)!i\binom{n+1}{i+1}}{n!} + \frac{k}{2}$
$= \sum_{i=k+1}^{n} \frac{k(i-2)!\cancel{(n-i)!}(n+1)!i}{n!\cancel{(n-i)!}(i+1)!} + \frac{k}{2}$
$= \sum_{i=k+1}^{n} \frac{k(i-2)!\cancel{n!}(n+1)i}{\cancel{n!}(i+1)!} + \frac{k}{2}$
$= k(n+1)\sum_{i=k+1}^{n} \frac{\cancel{(i-2)!}\cancel{i}}{\cancel{(i-2)!}(i-1)\cancel{i}(i+1)} + \frac{k}{2}$
$= k(n+1)\sum_{i=k+1}^{n} \frac{1}{(i-1)(i+1)} + \frac{k}{2}$

Iz Leme3: $\sum_{i=k+1}^{n} \frac{1}{(i-1)(i+1)} = \frac{-k^2(2n+1)+k(2n^2-1)+n(n+1)}{2k(k+1)n(n+1)}$
$= k(n+1) \sum_{i=k+1}^{n} \frac{1}{(i-1)(i+1)} + \frac{k}{2}$
$= \cancel{k}\cancel{(n+1)} \frac{-k^2(2n+1)+k(2n^2-1)+n(n+1)}{2\cancel{k}(k+1)n\cancel{(n+1)}} + \frac{k}{2}$
$= \frac{-k^2(2n+1)+k(2n^2-1)+n(n+1)}{2(k+1)n} + \frac{k}{2}$
S obzirom da tražimo k za fiksirano n to znači da je n konstanta.
Neka je:
$f(n,k) = -k^2(2n+1) + k(2n^2-1) + n(n+1)$
$\implies f'(n,k) = (-k^2(2n+1) + k(2n^2-1) + n(n+1))'$
$= (-k^2(2n+1))' + (k(2n^2-1))' + (n(n+1))' = -2k(2n+1) + 2n^2 - 1 + 0$
$= k(-4n-2) + 2n^2 - 1$
$g(n,k) = 2(k+1)n$
$\implies g'(n,k) = (2(k+1)n)' = \cancel{2'(k+1)n} + 2(k+1)'n + \cancel{2(k+1)n'} = 2 \cdot 1 \cdot n = 2n$
Tada je :
expValue$'$(n,k) $= (\frac{f(n,k)}{g(n,k)} + \frac{k}{2})' = \frac{f'(n,k)g(n,k)-f(n,k)g'(n,k)}{(g(n,k))^2} + \frac{1}{2}$
$= \frac{(k(-4n-2)+2n^2-1)2(k+1)n-(-k^2(2n+1)+k(2n^2-1)+n^2+n)2n}{4(k+1)^2n^2} + \frac{1}{2}$
$= \frac{(-4nk-2k+2n^2-1)(2kn+2n)-(-2k^2n-k^2+2n^2k-k+n^2+n)2n}{4(k+1)^2n^2} + \frac{1}{2}$
$= \frac{(-k^2\cdot 8n^2-k\cdot 8n^2-k^2\cdot 4n-k\cdot 4n+\cancel{k\cdot 4n^3}+4n^3-2kn-2n)-(-k^2\cdot 4n^2-k^2\cdot 2n+\cancel{k\cdot 4n^3}-k\cdot 2n+2n^3+2n^2)}{4(k+1)^2n^2} + \frac{1}{2}$
$= \frac{-k^2\cdot 4n^2-k^2\cdot 2n-k\cdot 8n^2-k\cdot 2n-k\cdot 2n+2n^3-2n^2-2n}{4(k+1)^2n^2} + \frac{n(k+1)^2}{2n(k+1)^2}$
$= \frac{-k^2\cdot 4n-k^2\cdot 2-k\cdot 8n-k\cdot 2-k\cdot 2+2n^2-2n-2}{4(k+1)^2n} + \frac{nk^2+2nk+n}{2n(k+1)^2}$
$= \frac{-k^2\cdot 2n-k^2-k\cdot 4n-k\cdot 2+n^2-\cancel{n}-1+nk^2+2nk+\cancel{n}}{2(k+1)^2n}$
$= -\frac{k^2\cdot n+k^2+k\cdot 2n+k\cdot 2-n^2+1}{2(k+1)^2n}$
$= -\frac{k^2(n+1)+2k(n+1)-(n-1)(n+1)}{2(k+1)^2n}$
$= -\frac{(n+1)(k^2+2k-n+1)}{2(k+1)^2n}$
$= -\frac{(n+1)((k+1)^2-n)}{2(k+1)^2n}$
Da bismo našli ekstreme funkcije trebamo izjednačiti njen izvod sa nulom.
expValue$'$(n,k) $= \delta$expValue$/\delta k = -\frac{(n+1)((k+1)^2-n)}{2(k+1)^2n} = 0$
Znamo: $k+1 > 0 \land n > 0$
$\implies (n+1)((k+1)^2-n) = 0$
$\implies (k+1)^2 - n = 0$
$\implies (k+1)^2 = n$
$\implies k+1 = \sqrt{n}$
$\implies \boxed{k = \sqrt{n} - 1}$
Znamo: δexpValue$^2/\delta k^2 = (-\frac{(n+1)((k+1)^2-n)}{2(k+1)^2n})'$
$= -\frac{((n+1)((k+1)^2-n))'2(k+1)^2n-(n+1)((k+1)^2-n)(2(k^2+2k+1)n)'}{(2(k+1)^2n)^2}$
$= -\frac{(n+1)((k+1)^2-n)'2(k+1)^2n-(n+1)((k+1)^2-n)2(k^2+2k+1)'n}{4(k+1)^4n^2}$

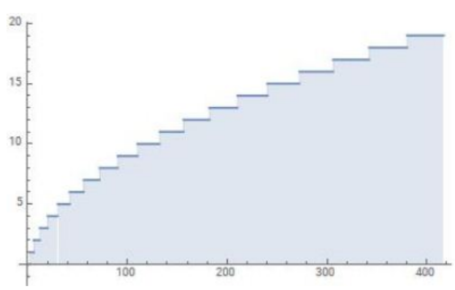
$$= -\frac{(n+1)(k^2+2k+1-n)'2(k+1)^2n-(n+1)((k+1)^2-n)2(2k+2)n}{4(k+1)^4n^2}$$
$$= -\frac{(n+1)2(k+1)2(k+1)^2n-(n+1)((k+1)^2-n)2\cdot 2(k+1)n}{4(k+1)^4n^2}$$
$$= -\frac{(n+1)\cancel{4}\cancel{(k+1)}\cancel{n}(\cancel{(k+1)^2}-\cancel{(k+1)^2}+n)}{\cancel{4}\cancel{(k+1)}(k+1)^3\cancel{n}\cdot n}$$
$$= -\frac{(n+1)\cancel{n}}{\cancel{n}(k+1)^3}$$
$$= -\frac{n+1}{(k+1)^3}$$

S obzirom da $k+1 > 0 \wedge n+1 > 0 \implies \delta \text{expValue}^2/\delta k^2 < 0$, što znači da je expValue(n,k) konkavna u tački $k = \sqrt{n}-1$, tj. $k = \sqrt{n}-1$ jeste maksimum te funkcije za posmatrane vrednosti. S obzirom da $k \in \mathbb{N}$ to znači da $\boxed{k = \lceil \sqrt{n}-1 \rceil \vee k = \lfloor \sqrt{n}-1 \rfloor}$

4 Grafici



Slika 1: Grafik zavisnosti očekivanog kvaliteta kandidata od ukupnog broja kandidata



Slika 2: Grafik zavisnosti broja kandidata koje odbijamo od ukupnog broja kandidata

5 Zaključak

U standardnom Sekretarovom problemu sekretar od n kandidata odbija prvih $e^{-1}n - 1$ dok u ovoj modifikaciji odbija prvih $\sqrt{n} - 1$.

Literatura

[1] F. Thomas Bruss (2000). Sum the odds to one and stop. Annals of Probability. 28 (3): 138491

[2] Optimal Stopping and Applications, retrieved on 21 June 2007

[3] Bearden, J.N., Murphy, R.O. Rapoport, A. (2005). A multi-attribute extension of the secretary problem: Theory and experiments. Journal of Mathematical Psychology. 49 (5): 410425

[4] F. Thomas Bruss (2003). A Note on Bounds for the Odds Theorem. Annals of Probability. 31: 18591961.